SPEECH SIGNAL RESAMPLING BY ARBITRARY RATE

ABSTRACT

In this paper we discussed issues related to resampling speech signal at arbitrary frequency by using interpolation methods. The implementation of four resampling methods, 1. direct interpolation, 2. Lagrange interpolation, 3. sine interpolation and, 4. Taylor series method, is presented. These methods have been tested with some speech data and various resampling frequencies. The quality of the resampled speech signals is analyzed and evaluated by human listening. The experiment results showed that either for upsampling and downsampling, sine and Lagrange methods generate additive high-frequency noise like metal sounds, but the direct and Taylor methods do not have some problem. The resampled speech by the direct and Taylor methods sounds more natural than that by sine and Lagrange methods.

1. INTRODUCTION

Continuous speech signal is usually converted into a series of discrete values by sampling at a proper frequency. The problem of resampling is to calculate the signal values at arbitrary time from a set of discrete signal samples since the original continuous signal is not available anymore. In general, the term “arbitrary time” is not so precise, it in fact refers to the “arbitrary time” within the original signal duration. From the viewpoint of mathematics, resampling is an interpolation problem; from the viewpoint of signal processing, resampling is a filtering process. The polynomial interpolation is the classical and natural method for resampling speech signal. In [1][2], the theory of bandlimited interpolation methods, i.e., Lagrange interpolation and sine interpolation, have been analyzed. In practice, the direct and Lagrange interpolation methods have been widely applied to speech signal with a lowpass filter.

In this paper we discussed four methods for resampling speech signal: 1. direct interpolation, 2. Lagrange interpolation, 3. sine interpolation and, 4. Taylor series representation. All of these methods can be used to resample speech signals by arbitrary sampling frequency. However, the resampled speech qualities are quite different by using these four methods. We compared the quality of the resampled speech signal by listening evaluation and signal-to-noise calculation. The experiments showed that for upsampling, sine interpolation and Taylor series representation are better than direct interpolation and Lagrange interpolation; for downsampling, Taylor series representation and direct interpolation are better than the other two methods.

2. INTERPOLATION

Suppose a speech signal x(n) (1<=n<=N) was sampled at frequency L, now we want to resample it at a new frequency M, where L and M are real numbers, M may be greater or less than L. If M>L, we call it upsampling; otherwise, downsampling. We denote y(k) (1<=k<=K) as the resampled values of x(n), and T_{old} = 1/L, T_{new} = 1/M are old and new sampling periods respectively. Obviously, the following relation should be satisfied:

$$K = (M/L)*N$$

Now, the object of the resampling problem is to compute y(k) (1<=k<=K) from x(n) (1<=n<=N).

To compute y(k) (1<=k<=K), some basic issues should be considered: (1). How many samples out of x(n) (1<=n<=N) are required to compute one sample of y(k) (1<=k<=K)? This refers to the window size problem. (2). What interpolation function is proper and how to estimate the approximation accuracy? This refers to the approximation function problem.

For the first problem, the speech signal property should be considered. We know that the short-term speech signal can be viewed as a stationary signal, but the long-term speech signal approximates random signal. Hence, we can assume that each sample of a speech signal only depends on some its neighboring samples. In practice, the window size is quite small, usually only contains a few samples, to save computation load and meet real-time requirement.

The interpolation function or approximation function is a mathematical problem. We know any continuous function can be arbitrarily and uniformly approximated by polynomials or triangle polynomials. The approximation accuracy can be estimated by using Taylor theorem and the original function's high order derivatives [2] (if the function has derivatives). Hence, in theory, we can choose proper polynomials or triangle polynomials as the interpolation function for speech resampling, i.e., Lagrange interpolation function is widely used in practice.

As the window size and interpolation function have been decided, the resampling performance has been consequently settled.

3. IMPLEMENTATION

Now we discuss the implementation of resampling a speech signal at arbitrary frequency. Suppose x(n) (1<=n<=N) is the original speech signal sampled at frequency L, y(k) (1<=k<=K) denotes the resampled values of the original signal at frequency M.

3.1 Direct Interpolation

The idea of direct interpolation is to approximate the points by using a line that is through two fixed points. For each time index k (1<=k<=K), let the real number n_k be:

$$n_k = (L/M)*k \qquad (3.1)$$

where L is the original sampling frequency and M is the new sampling frequency, L/M is the frequency scaling factor. For n_k, there must exist one time index n (1<=n<=N) which satisfies:

$$n <= n_k <= n+1 \qquad (3.2)$$

let the two real number weights w_1 and w_2 be:

$$w_1 = n_k - n, \quad \text{and} \quad w_2 = 1 - w_1$$

then the value of y(k) can be computed as:

$$y(k) = w_1*x(n+1) + w_2*x(n) \qquad (3.3)$$

Apparently, the computation of each y(k) only requires two original samples and y(k) is the weighted average of the two original samples.

3.2 Lagrange Interpolation

Since the sampling periods of speech signals are identical, it is natural to apply Lagrange interpolation method to speech resampling. The idea of Lagrange interpolation is to use polynomials to approximate the original continuous speech signal and to compute the resampling samples based on the Lagrange interpolation polynomials. The above direct interpolation method is a simplified special case of Lagrange interpolation method.

For each time index k (1<=k<=K), the real number n_k is computed as the above formula (3.1), and for n_k, suppose the time index n (1<=n<=N) satisfies condition (3.2). Suppose the highest degree of the Lagrange interpolation polynomials is 2*w, then the window size should be 2*w+1, the time index of the 2*w+1 original samples are as follows:

n-w, …, n-1, n, n+1, …, n+w

The value of y(k) can be computed by the following formula based on weighted average of 2*w+1 original samples.

$$y(k) = \sum_{i=-w}^{w} \frac{\prod_{j=-w}^{w} (n_k - (n - j))}{\prod_{j=-w}^{w} ((n - i) - (n - j))} x(n-i) \qquad (3.4)$$

where the subscript j of is from –w to w and not identical to i. Further, the denominator in the above (3.4) can be expressed as:

$$\Pi((n-i) - (n-j)) = \Pi(j-i) = (-1)^{w+i} * (w-i)! * (w+i)!$$

Where n! = 1*2* …*n and 0! = 1. Let $L_i(t)$ denote the function of the coefficients of above formula (3.4):

$$L_i(t) = \frac{\prod_{j=-w}^{w} (t - (n - j))}{\prod_{j=-w}^{w} ((n - i) - (n - j))}$$

It is known $L_i(t)$ converges to $\sin(\pi(t-i))/(\pi(t-i))$ while $w \to \infty$[2]. This means that the Lagrange interpolation approximates sine interpolation while the number of interpolation points is large.

3.3 Sine Interpolation

According to Shannon's sampling theorem the continuous speech signal x(t) can be reconstructed from the set of its discrete samples x(n) by the following formula:

$$x(t) = \sum_{n=0}^{N} \frac{\sin(\pi(t-n))}{\pi(t-n)} \qquad (3.5)$$

Therefore, the resampling is to compute new samples at arbitrary sampling frequency from the continuous form x(t). This is a D/A/D process.

The implementation is analogous to Lagrange method. For each time index k (1<=k<=K), the real number n_k is

computed as the above formula (3.1), and for n_k, suppose the time index n (1<=n<=N) satisfies condition (3.2). Suppose the window size is 2*w+1, the time index of the 2*w+1 original samples are as follows:

n-w, …, n-1, n, n+1, …, n+w

The value of y(k) can be computed by the following formula based on weighted average of 2*w+1 original samples.

$$y(k) = \sum_{i=-w}^{w} \frac{\sin(\pi(n_k - n + i))}{\pi(n_k - n + i)} x(n-i) \qquad (3.6)$$

By using formula (3.6), all samples of resampling process can be obtained.

3.4 Taylor Series Method

The continuous speech signal x(t) can be viewed as a time-domain smooth function. According to Taylor theorem, for any $t \in (t_i-\varepsilon, t_i+\varepsilon)$, x(t) can be expressed by an infinite series as follows:

$$x(t) = x(t_i) + (t- t_i)x'(t_i) + (t- t_i)^2 * x''(t_i)/2! + \ldots \quad (3.7)$$

Let T_{old} denote the original sampling period, and the t_is be the time index of the original sampling process,

$t_i = i * T_{old}$ and $\varepsilon = T_{old}/2$

then we can compute x(t) at arbitrary time index t based on some original samples only. Hence we can realize the resampling of speech signal x(t) at arbitrary sampling frequency.

The derivative and the second order derivative of x(t) at $t= t_i$ can be approximately obtained by the following numerical method.

$$x'(t_i) = (x(t_{i+1}) - x(t_i))/ T_{old}$$
$$x''(t_i) = (x'(t_{i+1}) - x'(t_i))/ T_{old}$$
$$= (x(t_{i+2}) - 2*x(t_{i+1}) + x(t_i))/ T_{old}^2$$

The computation of the new samples of resampling is as follows. For each time index k (1<=k<=K), the real number n_k is computed according to formula (3.1), and for n_k, suppose the time index n (1<=n<=N) satisfies condition (3.2), then we have:

$$y(k) = x(n) + (n_k - n)*x'(n) + (n_k - n)^2 * x''(n) \qquad (3.8)$$

where only the derivative and the second order derivative of x(t) are used, so three original samples x(n), x(n+1) and x(n+2) have been used to compute one new sample. If higher order derivatives being used, more original samples should be required and the approximation accuracy may be improved further.

3.5 Pseudo Code

x(n) (1<=n<=N) is the set of discrete samples of speech signal x(t) sampled at frequency L. M is the new frequency by which x(t) will be resampled. y(k) (1<=k<=K) denotes the new samples of x(t) resampled at frequency M.

```
(1). compute the frequency scaling factor:
       a = L/M
(2). for each time index k, 1<=k<=K
       find mapping index from k to the original
       time index n, 1<=n<=N.
            nk = (float) a*k ;
            n = (int) nk ;
       compute the value of new sample y(k)
       based on formula (3.3) or (3.4) or (3.6) or
       (3.8).
     end (for)
(3). low-pass filtering for post-resampling.
```

3.6 Post Resampling

Usually low-pass filtering is required for post resampling, especially when the resampling is a downsampling process. A FIR filter may be applied to the new samples obtained by the resampling process to cut off the frequencies which are beyond the bandlimit.

4. APPLICATIONS

The above four resampling methods have been tested with some speech data and various resampling frequencies. For Lagrange and Sine interpolation methods, the window length is 5 samples; for Taylor series method, the first and second order derivatives were used. The original speech signal was sampled at 8k Hz and the SNR is about 30db. The resampling frequencies are 3k, 6k, 10k, 14k Hz respectively. The resampled speech data without post-filtering were evaluated by human listening. The experiments showed that the resampled speech by Lagrange and sine interpolation methods have additive high frequency noise which likes metal sounds either upsampling or downsampling, but this problem does not exist in direct and Taylor method. By the listening evaluation results, the resampled speech quality of direct and Taylor method is better than Lagrange and sine interpolation methods. The SNR of the original and resampled speech were calculated and shown in Table 1.

Table 1. SNR of resampled speech

Rate	direct	Taylor	Lagrange	Sine
8k	30.25	30.25	30.25	30.25
3k	32.74	32.52	32.49	32.60
6k	31.22	31.23	31.66	33.20
10k	30.76	29.55	31.36	32.61
14k	30.74	29.15	32.08	31.98

Table 1 showed that SNR of speech signals does not change significantly by resampling process.

Based on the speech resampling experiment results, we can say the speech signal is local strong dependent. The dependency between samples will drop sharply while the distance between samples increases. This characteristic has been also shown in the speech autocorrelation function.

The following figures1-8 showed the original speech signal and the resampled speech signals by using different resampling methods respectively.

Figure 1. The original speech signal (8khz)

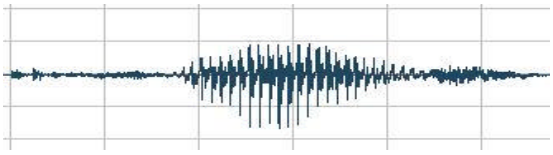


Figure 2. resampled by direct method(3khz)

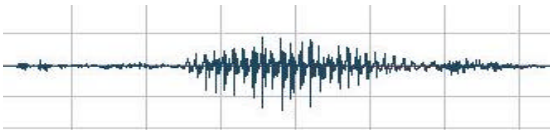


Figure 3. resampled by Taylor method (3khz)

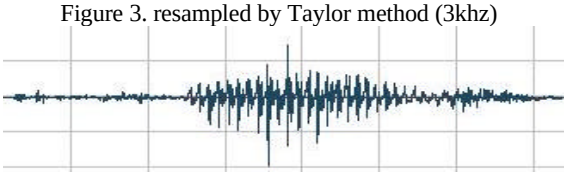


Figure 4. resampled by Lagrange method (3khz)

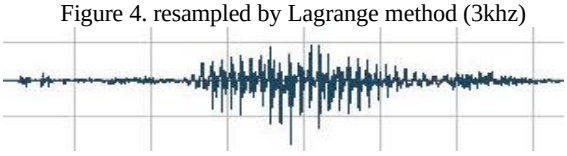


Figure 5. resampled by sine method (3khz)

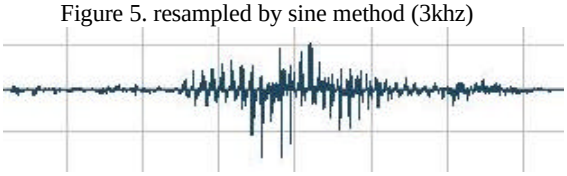


Figure 6. resampled by direct method (14khz)

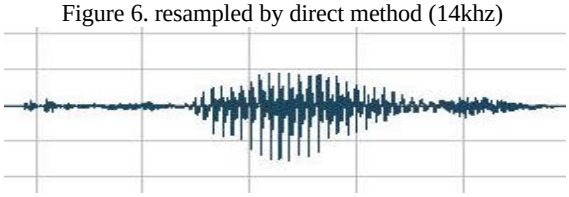


Figure 7. resampled by Taylor method (14khz)

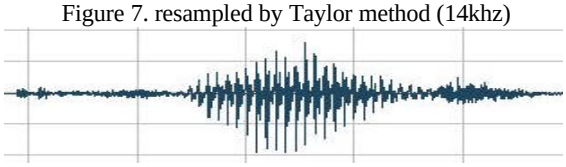
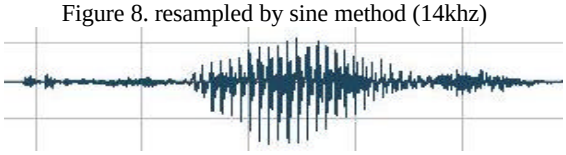


Figure 8. resampled by sine method (14khz)



From the above waveforms we can also find difference between the resampling methods, especially in downsampling, i.e., 8khz->3khz. The waveforms of the resampled signal obtained by sine and Lagrange methods changed significantly from the original one; the waveforms obtained by direct and Taylor methods are quite good.

We applied the resampling methods to speech synthesis to change the speech prosody. With upsampling the original speech signal, we can change the prosody lower, and downsampling can change the prosody higher. However, the original female voice will sounds like male voice by upsampling when the scaling factor is greater than 1.5; and male voice sounds like female voice by downsampling when the scaling factor is less than 0.5. These evaluations were conducted by human listening.

5. CONCLUSION

Four speech signal resampling methods and their implementations have been discussed which are convenient and efficient to convert digital speech signals from one sampling frequency to another one.

REFERENCES

[1] J.O.Smith, P.Gossett, Proc of ICASSP-84, Vol II, March 1984, New York

[2] http://www-ccrma.stanford.edu/~jos/resample